# Configuration Manual

MSc Internship
Programme Name

Eze Kenneth C.
Student ID: X19131178

School of Computing
National College of Ireland

Supervisor: Mr. Niall Heffernan

**National College of Ireland**

**MSc Project Submission Sheet**

**School of Computing**

| | |
|---|---|
| **Student Name:** | Eze Kenneth C. |
| **Student ID:** | X19131178 |
| **Programme:** | Cybersecurity **Year:** 2019/2020 |
| **Module:** | Internship |
| **Supervisor:** | Mr. Niall Heffernan |
| **Submission Due Date:** | 17th August, 2020 |
| **Project Title:** | CYBERCRIME DETECTION IN COMMUNICATIONS: AN EXPERIMENTAL CASE OF CYBER SEXUAL HARRASMENT ACCURACY DETECTION ON TWITTER USING SUPERVISED LEARNING CLASSIFIERS |

**Word Count:** ..... **Page Count:** .....

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

**Signature:** .....

**Date:** .....

**PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST**

| | |
|---|---|
| Attach a completed copy of this sheet to each project (including multiple copies) | □ |
| **Attach a Moodle submission receipt of the online project submission,** to each project (including multiple copies). | □ |
| **You must ensure that you retain a HARD COPY of the project**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. | □ |

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

| **Office Use Only** | |
|---|---|
| Signature: | |
| Date: | |

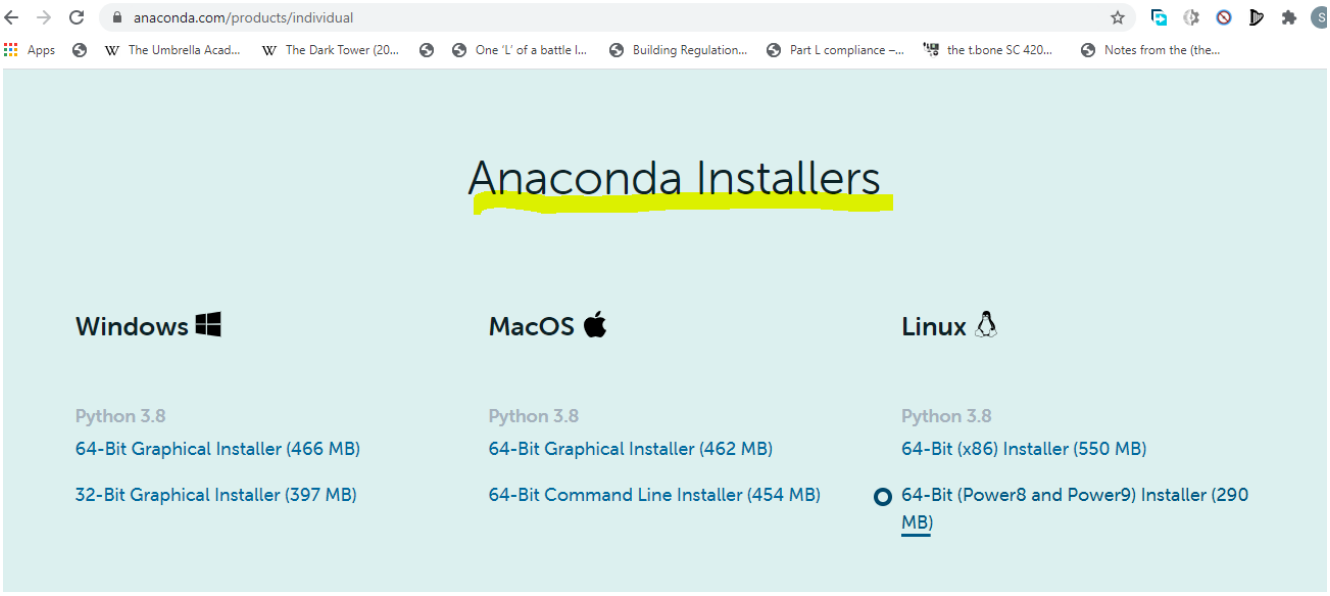
| | |
|---|---|
| Penalty Applied (if applicable): | |

•

Eze Kenneth C.
Student ID: X19131178

# 1 Download the Anaconda Installer Package

Download the anaconda installer package and run and install it

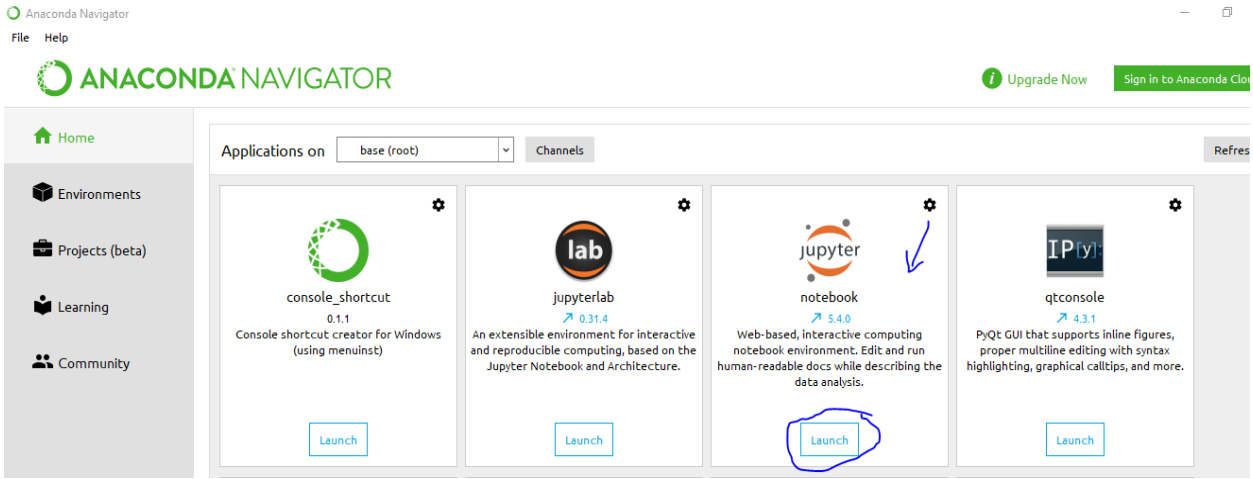



1.1 Install Required libraries in the python command prompt

Syntax: pip install <library name>

- Pip install tweepy
- Textblob
- Vadersentiment
- Pandas
- Xgboost
- sklearn


## 2 Now go into the Anaconda Navigation and launch the Jupyter Notebook




## 2.2 Locate the folder and click on the python program

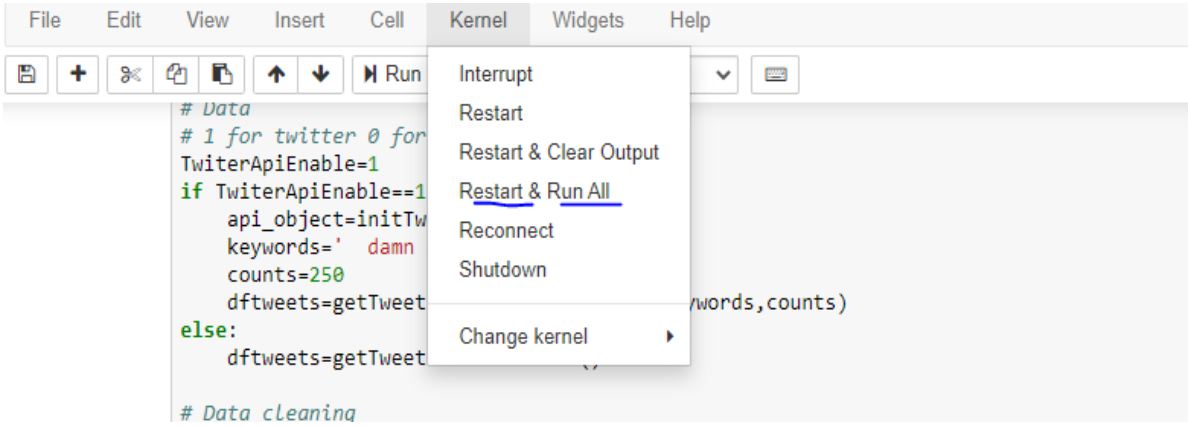
2.2 As we can see above, we have Kaggle csv file and python program. So, we can switch between the csv file in our testing and twitter api by enable either 1 for twitter api and 0.

Also, keywords for testing our accuracy

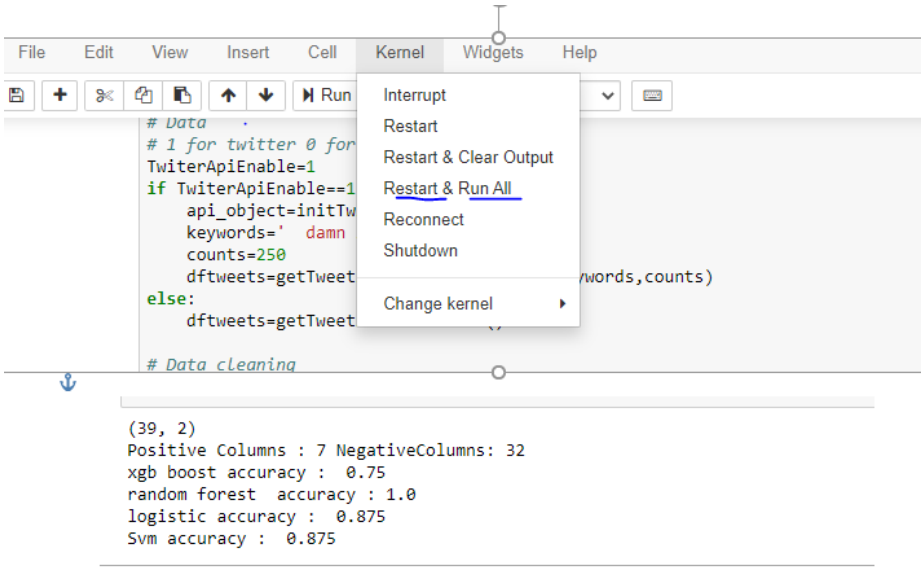
```
# program Stars
# Use this section to print result with data base
# Data
# 1 for twitter 0 for csv
TwiterApiEnable=1
if TwiterApiEnable==1:
    api_object=initTwitter()
    keywords='  damn stupid bitch like you'
    counts=250
    dftweets=getTweetsFromApi(api_object,keywords,counts)
else:
    dftweets=getTweetsFromDataBase()

# Data cleaning
dftweets=DataCleaning(dftweets)
#  Pos Tagging
dftweets=PostaggingStemming(dftweets)
# model accuracy
model(dftweets)
```

2.2 Now go onto the kernel and launch the python program



2.3 Below we can see the output of the program for further analysis




# 3 Mini Code Helper

The idea behind the code helper is to explain more to the reader some code functions they might find confusing in the code.

- **PostaggingStemming -** Using the nltk library we get the stop words here and clean the tweets. Also remove words which have characters less than 3 so that we only focus on words which are important
- **ParseTweet** - This function cleans the tweets and removes characters for ease of tokenization later.
- **DataCleaning -** This function is for cleaning the csv tweets.
- **getTweetsFromDataBase -** This function parses the tweets from the csv file and creates a data frame.
- **Model -** In this function we spilt the data in train and test and feed the train and test data in different models and get accuracy scores.